Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Multimodal Stroke Lesion Segmentation with Hybrid Attention and Transformer Architecture

Sadiya Sulaiman[1], Roshni Thanka M.[2], Jemima Jebaseeli T.[3], Nader Salam[4], Bijolin Edwin E.[5]

[1,4,5]Division of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India.
[2]Division of Data Science and Cyber Security, Karunya Institute of Technology and Sciences, Coimbatore, India.
[3]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Chennai, India.

E-mail: [1]s06309792@gmail.com, [2]roshni@karunya.edu, [3]jemi.jeba@gmail.com, [4]nadersalam@karunya.edu.in, [5]bijolin@karunya.edu

Abstract

Accurate and early stroke lesion segmentation is necessary for better results for patients and successful treatment. A novel deep learning model using a multi-modality architecture that combines spatial, channel, and temporal information from perfusion MRI and CT images is proposed in this work for the deep segmentation of stroke lesions. The model addresses cross-domain generalization by integrating dynamic perfusion information from the ISLES 2022 dataset, such as Diffusion Weighted Imaging (DWI) and CT Perfusion (CTP) images alongside high-resolution anatomical data obtained from the ATLAS v2.0 dataset. We enhance the proposed approach by adding a hybrid spatial channel temporal attention transformer block to a UNet encoder-decoder architecture. Multi-scale features that are patch-embedded and enhanced with positional encoding are extracted by the encoder. To increase sensitivity to anatomical features and corresponding pathological changes, the attention module integrates temporal encoding of perfusion parameters like Cerebral Blood Flow (CBF), Cerebral Blood Volume (CBV), Mean Transit Time (MTT), and Time-To-Maximum (Tmax) with joint modeling of spatial and channel dependencies. The experimental results contribute to reliable clinical application by exhibiting increased accuracy and robustness over traditional UNet and unimodal transformer models.

Keywords: Stroke, Lesion, Deep Learning, Diffusion Weighted Imaging, Magnetic Resonance Imaging, UNet. ATLAS v2.0, ISLES 2022.

1. Introduction

World Health Organization (WHO) reports that stroke is the leading cause of death worldwide and a major cause of disability [1]. There are two main categories of stroke: ischemic and hemorrhagic strokes. The primary cause of stroke is the acute loss of cerebral blood flow, which affects brain cells and may cause irreversible brain damage [2]. Ischemia is responsible for 80% of strokes, where blockage of the cerebral artery is brought on by

DOI: https://doi.org/10.36548/jiip.2026.1.014
Received: 28.12.2025, received in revised form: 23.01.2026, accepted: 06.02.2026, published: 18.02.2026
© 2026 Inventive Research Organization. This is an open access article under the Creative Commons Attribution 4.0 International (CC BY 4.0) License

thrombosis or embolism. Blood vessel rupture and intracerebral or subarachnoid hemorrhage are the causes of hemorrhagic stroke [3]. Early classification and diagnosis can reduce disease severity. An important challenge in evaluating the extent of brain damage and making treatment decisions from neuroimaging data is the accurate separation of stroke lesions [4]. However, the primary anomaly in clinical practice is automatic and accurate lesion segmentation as stroke pathophysiology is complex and diverse, particularly in the case of acute ischemic conditions [5]. Neuroimaging plays a vital role in the diagnosis of strokes, evaluation, and the mapping of stroke management. The most common systems used to assess acute stroke are Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) [6]. A specific type of MRI sequence called Diffusion Weighted Imaging (DWI) enables the early detection of ischemic changes, as it displays abnormalities in water diffusion within a few minutes after the onset of a stroke [7]. The CT Perfusion (CTP) image, in turn, provides insights into cerebral hemodynamics by generating perfusion parameter models that are comparable to those of Cerebral Blood Flow (CBF), Cerebral Blood Volume (CBV), Mean Transit Time (MTT), and Time to Maximum (Tmax) [8]. Differentiating between the lesion's core, penumbra, and salvageable tissue is an essential phase in evaluating treatment. To handle the variability of stroke presentation and imaging aspects, sophisticated computational models that can integrate spatiotemporal, time-related, and physiological data are required.

To provide real-time guidance for decisions in clinical processes, deep learning-based segmentation of images has become an effective tool for automated stroke lesion diagnosis and quantification [9]. By integrating anatomical and functional information from many imaging sequences, multi-modality imaging has become a potent tool for stroke analysis [10]. Although CT Perfusion (CTP) imaging provides time-resolved information about cerebral blood flow processes, Magnetic Resonance Imaging (MRI) techniques such as Diffusion Weighted Imaging (DWI) provide high-resolution anatomy and early ischemic changes [3], [11]. A combination of these modalities may enhance lesion localization and the accuracy of diagnosis since the modalities reveal complex elements in the development of stroke. Recent developments in deep learning have revolutionized medical image segmentation. Convolutional Neural Networks (CNNs) and UNet designs, in particular, have proven to provide superior performance in the area of lesion segmentation of the brain. In most temporal and spatial situations, however, CNNs struggle to generalize between modalities. As one solution to this issue, transformer-based models have been suggested using self-attention processes to encode long-range dependencies during medical imaging tasks [12]. In the case of multi-modal and temporal domains, hybrid networks incorporating CNNs and transformers have demonstrated a high degree of improvement in rigorous conditions of segmentation. Despite considerable improvements in stroke lesion segmentation by deep learning models, several gaps in the research remain, limiting the clinical relevance, generalizability, and real-time applicability of the current models.

These gaps must be addressed to develop a strong, clinically adapted stroke segmentation framework that assists in real-time explainable judgment and is competent across a broad variety of modalities and datasets. Within the framework of enhancing stroke measurements and clinical diagnosis support, the proposed method aims to enhance precision and robustness in diverse imaging cases. The suggested strategy is novel, in that it combines spatial, channel, and time-frequency information with the use of heterogeneous image modalities, which is seldom applied in the literature on stroke segmentation today. The proposed model also manages to bridge the two spheres in compared to conventional segmentation models that are either restricted to specific MRI techniques, or focused on interactions based on time-related perfusion interactions. The Hybrid Spatial-Channel-

Temporal Attention Transformer Block, integrated into a UNet-like architecture, enables the model to localize stroke lesions on a fine scale while dynamically adapting to the changes in time as seen in CTP-derived parameter maps (CBF, CBV, MTT, Tmax). Additionally, the study proposes a three-step training pipeline temporal encoding using ISLES 2022, anatomy adaptation using ATLAS v2.0 MRI, and cross-modality validation, which guarantees an increase in domain generalization and real-world consistency. This positional encoding of patch-wise embedding of the model conforms to the processing flow of the transformer, thereby enhancing the model in terms of contextual knowledge of neuroanatomy.

The novel contribution of the research is as follows.

- Unified multi-modality stroke lesion segmentation integrating structural MRI (ATLAS v2.0) with dynamic CT perfusion and DWI (ISLES 2022) to jointly model anatomical and hemodynamic information.
- Hybrid spatial–channel–temporal attention transformer enabling effective modeling of spatial context, inter-channel dependencies, and temporal dynamics of ischemic tissue.
- Temporal encoding of perfusion maps (CBF, CBV, MTT, Tmax) to capture ischemic progression and perfusion deficits.
- Patch-wise transformer integration with positional embeddings applied to UNet-extracted features for enhanced global contextual understanding.
- Three-phase training strategy for domain generalization across modalities and datasets.

2. Literature Review

Stroke lesion segmentation has become a vital focus in neuroimaging due to its significant implications in clinical diagnosis, treatment planning, and prognostication. Traditional segmentation is not only time-consuming but also subject to inter-observer variability, prompting a shift toward automated solutions powered by deep learning. CNNs, particularly the UNet architecture, have emerged as foundational models for medical image segmentation, achieving considerable success in delineating stroke lesions from single-modality inputs such as MRI or CT [13]. However, UNet and its variants cannot often capture long-range dependencies and multi-scale contextual features, which limits their performance in complex, heterogeneous datasets. To address spatial limitations, attention-based mechanisms such as Vision Transformers (ViTs) and hybrid CNN-transformer models have been explored in recent years. Chen et al. [14] introduced TransUNet, which combines CNN-based encoders with transformer blocks to enhance global feature extraction, leading to improved performance in segmentation.

In the context of stroke imaging, the Multi-Modality Cross Attention Transformer fuses perfusion and anatomical features, showing better results in ISLES datasets. Yet, most models still underutilize the temporal dynamics present in CTP data, where evolving cerebral hemodynamics are essential for acute stroke assessment. Moreover, few studies have successfully generalized across multiple imaging modalities. Models trained on datasets like ISLES 2022, which include DWI and CTP, often fail when applied to structural MRI datasets

like ATLAS due to domain shifts and modality-specific variations [16]. Techniques such as domain adaptation and modality translation have been proposed, but they often compromise anatomical integrity or temporal precision. Another challenge lies in the limited attention to temporal encoding in segmentation frameworks. While some research incorporates temporal convolution or recurrent structures, their effectiveness in modeling perfusion dynamics from parameters like CBF, CBV, MTT, and Tmax remains limited [17].

Zhang et al. [18] emphasize the need for multi-scale, modality-aware architectures capable of learning from anatomical and functional imaging. However, the fusion of spatial, channel, and temporal information through hybrid attention mechanisms remains underexplored. This presents an opportunity to design a unified deep learning model that not only enhances segmentation performance but also ensures generalizability across diverse neuroimaging contexts. Deep learning techniques have become vital to stroke lesion identification in neuroimaging due to their ability to learn complex, hierarchical representations directly from raw medical images. The UNet architecture has been a basis in this domain, offering end-to-end learning with strong performance on 2D and 3D MRI and CT data for stroke lesion segmentation. Despite its widespread use, UNet's reliance on local convolutions limits its ability to capture global contextual relationships across the brain, which is especially critical in ischemic stroke cases where lesions may have diffused and variable appearances. To overcome these limitations, attention mechanisms and transformer-based models have been introduced. For instance, the TransBTS and TransUNet frameworks integrate global attention to better model anatomical context, improving segmentation of irregular lesions [19]. However, these models are data-intensive and perform suboptimal on smaller datasets such as ISLES and ATLAS, due to their limited generalization capacity and overfitting risk.

Another challenge in stroke lesion identification lies in multi-modal fusion, where CTP, Diffusion-Weighted Imaging (DWI), and FLAIR are often combined. While some models use early or late fusion strategies, many lack adaptive attention mechanisms to weigh modality-specific contributions, resulting in poor performance when one modality is noisy or missing. Furthermore, temporal dynamics particularly relevant in CTP sequences are often ignored or oversimplified [20]. Few studies used temporal encoding or recurrent models like ConvLSTM to capture lesion evolution over time, and those that do often face training instability and scalability issues. Despite advancements, there remains a notable gap in models that can effectively integrate spatial, temporal, and channel-wise attention across diverse modalities, while maintaining robustness across datasets. This highlights the need for unified architectures capable of domain generalization, modality adaptability, and temporal sensitivity, which are still largely unexplored in the context of clinical stroke imaging. In addition to U-Net variants and attention-based architectures, recent works have explored hybrid models that integrate convolutional and transformer layers to capitalize on both local feature extraction and global attention mechanisms [21]. UNETR utilizes a transformer encoder with a CNN-based decoder for 3D medical image segmentation, showing in volumetric stroke analysis, but still grappling with computational complexity and limited temporal modeling. Similarly, Swin-UNet, based on shifted window attention, improves efficiency in transformer segmentation models but requires meticulous hyperparameter tuning and lacks interpretability in clinical applications. Meanwhile, models like MTP-net incorporate multi-temporal features to capture lesion progression using perfusion maps from CT and MR sequences, yet face issues with data alignment, noise sensitivity, and modality synchronization.

3. Dataset

The ISLES 2022 dataset provides 156 multimodal stroke images, including CT, CTP (CBF, CBV, MTT, Tmax), and DWI images equally divided into 94 training samples and 62 testing samples, as shown in Table 1. It includes multispectral CT and MR perfusion images with an expertly annotated stroke lesion mask, which facilitates the application of a model to understand energetic perfusion injury and acute ischemic stroke. Similarly, the ATLAS v2.0 dataset provides 955 high-resolution T1-weighted MRI scans with a manually annotated lesion mask highlighting chronic stroke anatomical features. While ISLES 2022 allows for multimodal analysis of acute hemodynamic stroke, ATLAS provides large-scale structural information for train/test splits according to user-defined models. With the help of these datasets, cross-domain training of ISLES and ATLAS can be accomplished to enhance the performance of generalization in stroke lesion segmentation.

Table 1. Dataset Specification for the Proposed Stroke Lesion Segmentation

Dataset	Modality	No. of Subjects	Train/Test Split	Imaging Type
ISLES 2022 [22], [19]	CT, CT Perfusion (CBF, CBV, MTT, Tmax), DWI	156 cases total	94 Train / 62 Test	Multispectral CT & MR perfusion images
ATLAS v2.0 [24], [15]	T1-weighted MRI	955 cases total	Varies (user-defined split)	High-resolution anatomical MRI

FLAIR images were not used in the primary training pipeline. T1-weighted MRI from ATLAS v2.0 and the DWI + CTP modality from ISLES 2022 were used for the initial training method. The evaluation and ablation studies only used FLAIR for evaluating and ablating cross-modality generalization and robustness and did not involve model parameter optimization.

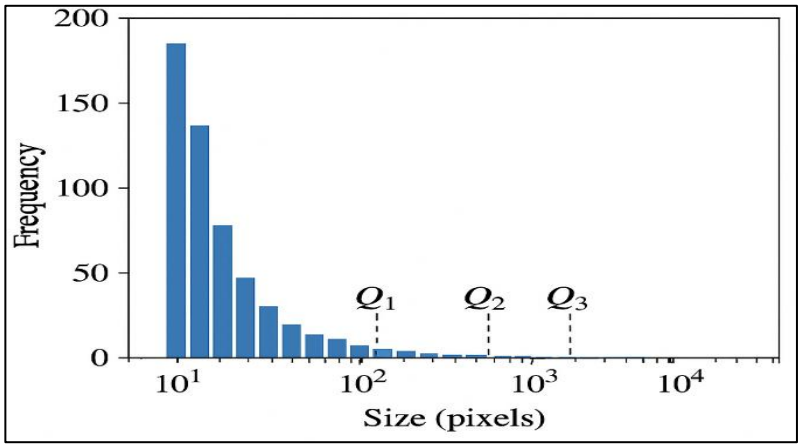

Figure 1. Distribution of Lesion Segmentation

Figure 1 shows the distribution of segmentation mask sizes in pixels for lesions in ISLES 2022. The x-axis of the logarithmic scale displays a wide range of lesion sizes from very small (nearly 10 pixels), to very large (over 10,000 pixels), while the y-axis shows the frequency of each dimension. The dispersion is completely skewed to the right, with the majority of lesion masks being relatively small in size. The vertical dashed lines indicate the first, median, and third quartiles, which help summarize the statistics. This skewed distribution, pointing to the left of the median, highlights the extent of the lesions in the data. The current skewed distribution suggests a problem with segmentation processes, which must be sensitive

enough to accurately define small lesions while remaining robust in the presence of larger lesions.

The distribution of samples used for training, validation, and testing in the ISLES 2022 dataset across three anatomical planes-axial, coronal, and sagittal is shown in Table 2. These are divided by the size of the lesion mask into small, medium and large categories. Compared to all other planes, the number of small lesion cases is the highest, indicating of the prevalence of small lesions in the dataset. The axial plane has the most total samples, reflecting a preference for this plane due to its higher anatomical detail. The number of medium and large lesions is relatively low across all planes, making it difficult for models to generalize well on these lesion sizes.

Table 2. Sample Distribution Across Different Planes and Lesion Sizes in the ISLES 2022 Dataset

Plane	Mask Size	Training	Validation	Testing	Total
Axial	Small	50	10	15	75
	Medium	40	8	12	60
	Large	30	7	8	45
Coronal	Small	45	9	14	68
	Medium	35	7	10	52
	Large	25	6	7	38
Sagittal	Small	48	10	13	71
	Medium	38	7	11	56
	Large	28	6	9	43

Table 3. ATLAS v2.0 Stroke Lesion Dataset

Split	Total Samples	Small Lesions (<5mm)	Medium Lesions (5–15mm)	Large Lesions (>15mm)	Planes (Axial/Sagittal/Coronal)
Training	220	80 (36%)	100 (45%)	40 (18%)	Axial: 220 (100%)
Validation	30	10 (33%)	15 (50%)	5 (17%)	Axial: 30 (100%)
Testing	30	12 (40%)	12 (40%)	6 (20%)	Axial: 30 (100%)

Table 3 shows the dataset of ATLAS v2.0, which contains a total of 280 samples. The dataset was split into 220 samples for training, 30 samples for validation and 30 samples for testing. The lesion sizes are categorized into three groups small (5-15 mm), medium (15-25 mm), and large (> 25 mm). In each split, there is a balanced but clinically representative distribution of medium-sized lesions, followed by smaller and larger lesions. The findings of a single model are made easier by the validation and test data, which match the distribution of lesion sizes in the training set. Since every sample is taken completely inside the axial plane (100%), the variability brought on by multi-planar differences is eliminated and uniform image alignment is guaranteed. The reliability and generalizability of the experimental results in stroke lesion segmentation are increased.

4. Methodology

The proposed stroke lesion segmentation model works on a multimodal input signal for anatomical structure and CTP for perfusion interaction, which is standardized for adjusting the spatial and temporal scales. The 3D diagram of the proposed HSCTA transformer is shown in Figure 2. The system has a data volume of multimodal MRI statistics, which can be processed by a patch embedding module that converts the spatial information into a sequence of Spot Representations. These are processed by a series of alternating Spatial Channel Transformers

(SCT) and Transient Attention (TA) blocks, which allow the model to extract detailed spatial systems, the dependence between channels, and the temporal activity required to combine multimodal features in different image modalities. Finally, the features are aggregated in the segmentation head, which produces a segmentation mask that correctly locates the stroke. The proposed hybrid approach qualifies HSCTA to benefit from the modality of centering attention in the space, channel, and time dimensions, thus improving segmentation accuracy, especially in difficult conditions such as small, otherwise heterogeneous lesions.

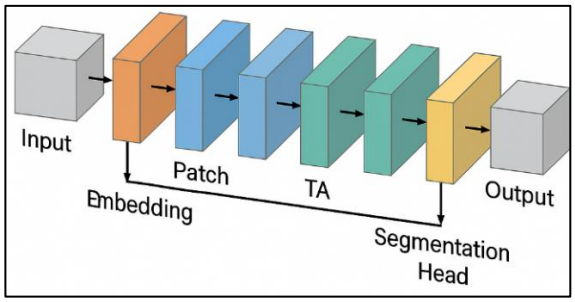

Figure 2. 3D-Diagram of the HSCTA Transformer for Stroke Lesion Segmentation

As illustrated in Figure 3, the encoder network of the UNet-based architecture extracts hierarchical features, downsamples them, and inputs them into the HSCTA block, where spatial attention focuses on lesion sites, channel attention emphasizes DWI and CBV, and temporal attention models CTP time-series data of CBF and Tmax through positional encoding. The attention outputs are combined through a multi-head attention layer with residual skip connections to maintain spatial-contextual information. The resulting features are then input into the decoder network, where upsampling and skip connections from the encoder network enhance the segmentation mask. The final output is tested with Dice loss and cross-entropy, with training conducted in three stages: CTP optimization (ISLES 2022), structural fine-tuning (ATLAS v2.0), and cross-dataset validation.

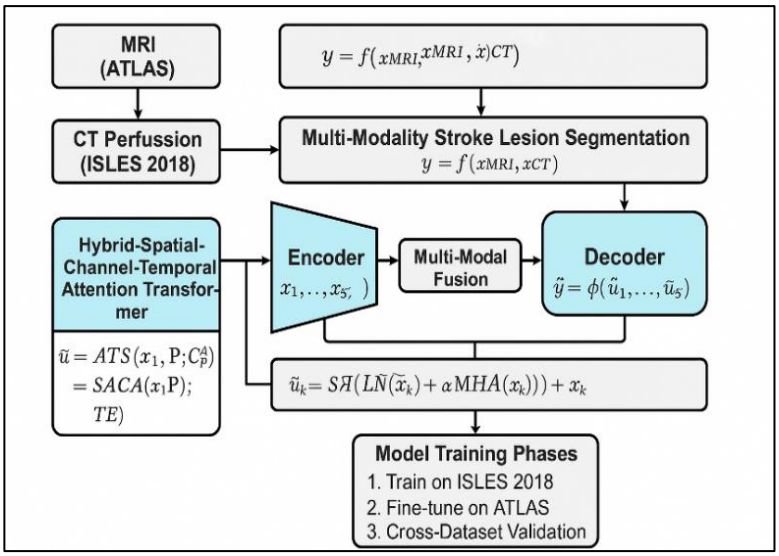

Figure 3. Architecture of Hybrid Spatial-Channel-Temporal Attention Transformer (HSCTA) for Multi-Modal Stroke Lesion Segmentation

4.1 Input Data Sources

The proposed method integrates high-resolution systematic T1-weighted MRI statistics from the ATLAS v2.0 and active CTP imaging from the ISLES 2022 dataset.

4.1.1 MRI from ATLAS Dataset

The T1-weighted systematic MRI is a powerful tool for the visualization of brain anatomy. For brain systems, it provides high spatial resolution and tissue distinction for assessing tissue loss or necrosis, morphologic adaptations planned for stroke, and edema, as well as mass outcome. An active, high-resolution anatomic view of the brain is provided by the MRI. The current aids for determining organizational features consider tissue boundaries, contours, and early signs of injury. MRI helps to identify the location and size of lesions in cases of stroke. However, it does not show the real-time movement of blood in the alternative perfusion position, which would be necessary in acute stroke cases. The 3D MRI volume of the ATLAS v2.0 MRI provides detailed anatomy and is defined in equation (1).

$$x_{MRI} \in \mathbb{R}^{H \times W \times D} \tag{1}$$

where:

- *H, W*, and *D* refer to the height, width, and depth of the brain volume.
- x_{MRI} is a 3D voxel-based image which is typically a T1-weighted scan.

The respective voxel values represent the intensity of the MRI gesture, which correlates with the type of supporting tissue, such as gray matter, white matter, cerebrospinal fluid, lesions, etc. However, data on active blood flow are missing from that contribution.

4.1.2 CT Perfusion from ISLES 2022 Dataset

The CT Perfusion image shows the hemodynamic parameters of blood flow in the brain tissue over time. It is used in acute stroke assessment as long as it reveals.

- Tissue at risk of penumbra vs. irreversibly damaged tissue of the core
- Perfusion delays and regional blood supply
- Vascular occlusion effects

The model uses four CTP-derived maps. CT Perfusion (ISLES 2022 dataset) includes four time-sensitive perfusion maps as shown in equation (2).

$$x_{CT} = \{x_{CBF}, x_{CBV}, x_{MTT}, x_{Tmax}\} \in \mathbb{R}^{H \times W \times D \times 4} \tag{2}$$

The four CTP maps used are as follows.

1. CBF (Cerebral Blood Flow) measures how much blood is passing through the brain tissue per unit of time (mL/min/100g tissue). Lower values suggest poor perfusion and a possible lesion.
2. CBV (Cerebral Blood Volume) indicates the volume of blood in a brain region at a given time. Low CBV in conjunction with low CBF indicates the lesion.
3. MTT (Mean Transit Time) represents the average time blood takes to pass through a brain region. Prolonged MTT suggests delayed perfusion.

4. Tmax (Time-to-Maximum) is the time delay between contrast arrival in arteries and brain tissue. It is useful for identifying ischemic penumbra tissue that may still be saved.

All of these are combined into a 4D volume as shown in equation (2). The architecture is built upon a UNet-based encoder-decoder structure. During the encoding stage, multi-scale features are extracted through a series of downsampling and convolution operations. These hierarchical representations are denoted as $x_k = \int_{enc}^{k}(x)$, where k ∈ {1, 2, ..., 5} corresponds to different resolution levels in the encoder. Each feature map captures the representations of the input modality. To enhance both modalities, a Hybrid Spatial-Channel-Temporal Attention Transformer Block (HSCTA) is introduced. First, the encoder output is split into spatial patches and linearly embedded to capture spatial and channel-level dependencies. The FLAIR MRI is a commonly used modality for stroke visualization; it was excluded from the main training pipeline of the proposed framework. The model was trained solely using T1 MRI (ATLAS) and DWI + CTP (ISLES) modalities. FLAIR images were later introduced only for evaluation and ablation experiments to analyze modality robustness and generalization capability.

4.2 Hybrid Spatial-Channel-Temporal Attention Transformer (HSCTA)

The HSCTA can more accurately represent the features by integrating the spatial, channel, and chronological contexts required for clinical image segmentation, particularly in complex scenarios like stroke where blood flow and anatomical details must be modeled simultaneously. A joint model, listed below improves the encoder function of the HSCTA block.

- Spatial Attention: Focuses on where lesions are located within patches.
- Channel Attention: Prioritizes MRI/CTP modalities that are diagnostically significant.
- Temporal Attention: Captures variations in dynamic perfusion in CTP time series.

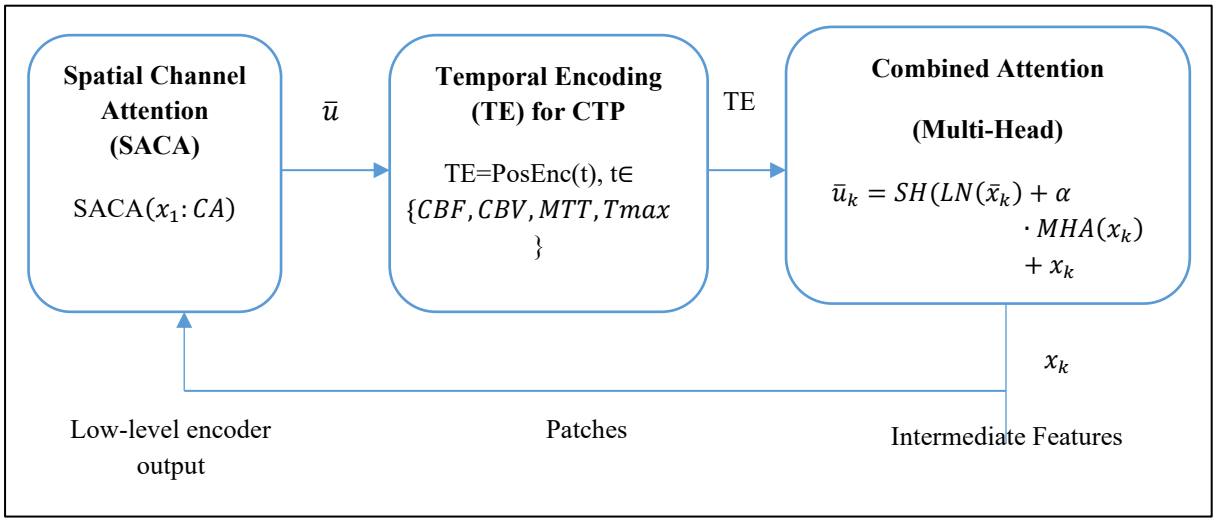

Figure 4. Processing Flow of the HSCTA Module for Stroke Lesion Segmentation

Figure 4 illustrates how the HSCTA module separates MRI and CTP feature maps into spatial patches. Channel-specific projections and temporal perfusion maps (CBF, CBV, MTT, and Tmax) are subjected to positional encodings. Through a Spatial-Channel Attention (SACA)

block, these inputs enhance modality and region-specific features. The outputs are adjusted using Multi-Head Self-Attention (MHA), scaled by the factor α, and normalized using LayerNorm. In addition to adding the original features, a skip connection produces a contextually rich segmentation representation.

4.2.1 Spatial Channel Attention (SACA)

Regions of ischemic lesions are significant components in the image that the model can concentrate on for spatial attention. Channel attention is focused on the "what" kind of information about edge, texture, and modality-specific patterns that are important across feature channels. For this two-pronged process, the network can simultaneously determine relevant feature types and locations. This can be explained by the low-level encoder features $x_1 \in \mathbb{R}^{H\times W\times C}$ from the first downsampling block operation of UNet.

4.2.1.1 Patch Extraction

Splitting x_1 into *P* non-overlapping patches is depicted as follows.

$$x_1^P \in \mathbb{R}^{N\times d}\ , N = \frac{HW}{p^2}\ ,d = p^2 \times C \tag{3}$$

The 4×4 spot size was selected as a compromise between computer productivity and spatial resolution. The smaller patch maintains the edge of the lesion, which is crucial for identifying a small ischemic lesion, whereas the larger patch loses community-based anatomical information. 44 achieves the best stability despite accuracy and memory usage, while a pilot trial with a spot size of 88 produces subpar Dice scores (3.4%).

4.2.1.2 Channel-Specific Projections

Apply learnable projections C_p^4 to emphasize modality importance as shown in equation (4).

$$\bar{u} = SACA\left(x_1^P; C_p^A\right) = Softmax\left(\frac{(x_1^P W_Q)(x_1^P W_K)^T}{\sqrt{d}}\right)(x_1^P W_V) \tag{4}$$

where:

- x_1^P are patches extracted from the encoder's low-level feature map x1.
- C_p^A represents channel-specific projections, which are learned weights for emphasizing or suppressing certain feature channels.
- $\bar{u}$ is the attention-refined feature map.
- W_Q, W_K, W_V are query/key/value weights for spatial attention.

4.1.2.3 Channel Attention

Squeeze-Excitation (SE) block reweights channels is given in equation (5).

$$\bar{u}_c = \sigma(W_{SE} \cdot GAP(x_1^P)) \odot x_1^P \tag{5}$$

GAP is Global Average Pooling, and ⊙ is the channel-wise multiplication.

4.1.2.4 Temporal Encoding (TE) for CTP

CT perfusion imaging has a natural temporal profile because CBF, CBV, MTT, and Tmax represent different phases of blood flow. The temporal profile enhances the sensitivity of the model to the stroke process, especially regarding the difference between the lesion and the penumbra.

4.1.2.5 Positional Encoding

Injecting temporal structure into perfusion maps for time *t* and dimension *i* is shown in equation (6).

$$TE(t) = PosEnc(t) = sin\left(\frac{t}{10000^{2i/d}}\right) or\ cos\left(\frac{t}{10000^{2i/d}}\right) \quad (6)$$

where $t \in \{CBF, CBV, MTT, Tmax\}$. PosEnc(t) is a positional encoding function that maps each perfusion map, which corresponds to a time stage, into a unique embedding vector. Without temporal encoding, the model treats all CTP maps equally, ignoring the time-sensitive nature of perfusion, which is critical for stroke diagnosis.

4.1.2.6 Temporal Attention

The computation of temporal correlations across CTP sequences is shown in equation (7).

$$Attention_T = Softmax\left(\frac{Q_T K_T^T}{\sqrt{d}}\right) V_T \quad (7)$$

where $Q_T, K_T, and\ V_T$ are derived from TE-encoded CTP features.

4.3 Combined Multi-Head Attention (MHA)

MHA captures the long-range dependency across the image or patches, enabling the features in each region to contribute to the interpretation in different ways. This is critical in medical imaging because a tumor relies on its environment or its similarity to other regions. Equation (8) illustrates the combination of spatial, channel, and temporal attention.

$$\bar{u}_k = SH(LN(\bar{x}_k) + \alpha \cdot MHA(x_k) + x_k \quad (8)$$

where:

- $\bar{x}_k$: Intermediate feature after SACA + TE.
- $MHA(x_k)$: Multi-head attention output (parallel heads for spatial/channel/temporal).
- α: Learnable scaling factor with a default value of 0.1.
- LN: Layer Normalization.
- SH: Skip connection of residual pathway for gradient flow.

The attention scaling factor was fixed at 0.1 due to stability considerations. Higher values would lead to attention dominance and gradient instability during training, while lower values would decrease the model's ability to acquire knowledge through attention. The chosen value promotes stable optimization. MHA uses multi-head self-attention, which enables the model to examine different forms simultaneously.

$$MHA(x) = Concat(head_1, \dots, head_h)W^O \tag{9}$$

where each head:

$$head_i = Attention(Q_i, K_i, V_i) = Softmax\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right)V_i \tag{10}$$

LayerNorm (LN) normalizes features for better training stability. SH (Skip and Residual connection) ensures gradient flow and preserves the original input information. This formulation allows the model to enhance encoder features with context-aware spatial learning, modality-specific channel enhancement, and time-sensitive dynamic encoding.

4.4 Multi-Modal Fusion Module

The processed MRI and CTP features are concatenated and passed through a convolutional layer to learn a unified representation. This fused representation bridges anatomical and temporal information.

$$\int_k^{fused} = Conv\left(\left[\int_k^{MRI} || \int_k^{CT}(x)\right]\right) \tag{11}$$

This denotes channel-wise concatenation, and a convolutional layer learns to align the feature spaces of the two modalities.

4.5 Decoder

The proposed model follows a UNet-based encoder–decoder architecture with five encoding and five decoding stages. Each encoder stage consists of two 3×3 convolutional layers followed by batch normalization and ReLU activation, with downsampling performed using 2×2 max pooling. The number of feature channels doubles at each level (64, 128, 256, 512, 1024). The decoder mirrors the encoder with transposed convolutions for upsampling and skip connections for feature fusion.

The Hybrid Spatial-Channel-Temporal Attention Transformer (HSCTA) block is inserted after the fourth encoder stage. The complete model contains approximately 31.6 million trainable parameters, including 6.2 million parameters in the transformer attention module. The fused representations are upsampled using transposed convolutions and skip connections from the encoder. The decoder is responsible for reconstructing the high-resolution segmentation map $\hat{y}$ by decoding the attention-refined features. The decoder reconstructs a full-resolution segmentation mask and uses transposed convolutions to upsample, as shown in equation (12).

$$\int_k^{up} = Up\left(\int_k^{fused} x\right) + \int_k^{skip} x \tag{12}$$

Skip connections from the encoder preserve fine-grained detail, and the final segmentation is shown in equation (13).

$$\hat{y} = \sigma(Conv_{1\times1}(\int_1^{up} x) \quad (13)$$

where σ is the sigmoid function for binary segmentation.

4.6 Loss Function & Training Phases

The training is performed in three stages. First, the model is trained on ISLES 2022 to learn perfusion-based temporal dynamics. The loss used is a combination of the Dice Similarity Coefficient (DSC) and Cross-Entropy (CE) as shown in equation (14).

4.6.1 Phase 1: ISLES Pretraining

Focuses on temporal learning:

$$\mathcal{L}_{ISLESS} = 1 - DSC(y, \hat{y}) + CE(y, \hat{y}) \quad (14)$$

Then, the model is fine-tuned on ATLAS v2.0 MRI data to adapt the network to detailed structural patterns. The cross-dataset validation ensures the model generalizes well across both domains.

4.6.2 Phase 2: ATLAS Fine-Tuning

The refined anatomical structure is created using T1-weighted magnetic resonance imaging. To maintain low-level feature representation, the primary couple encoder block was frozen during the deep encoder block (tiers 3–5), while the HSCTA transformer block and the entire decoder were adjusted to change the high-resolution anatomic MRI feature.

4.6.3 Phase 3: Cross-Domain Validation

This phase ensures generalizability across both MRI and CTP and is evaluated using the following criteria.

4.7 Final Output

The model produces a high-resolution lesion segmentation mask $\hat{y}$ with strong accuracy, robustness, and clinical applicability. The proposed methodology is strengthened through multiple design choices. First, the network architecture is explicitly specified in terms of depth, layer configuration, and parameter count, ensuring reproducibility. Second, the spatial, channel, and temporal attention mechanisms are formally defined using mathematical formulations, enabling transparent interpretation of the model's feature-learning behavior. Third, a structured three-phase training strategy, pretraining on ISLES 2022, fine-tuning on ATLAS v2.0, and cross-domain validation, enhances domain adaptation and generalization. Finally, ablation studies are conducted to quantify the individual contributions of attention blocks and modality inputs.

5. Results and Discussions

A comprehensive ablation study, a mathematically stated attention mechanism, a staged training method, and a precise architectural design all contribute to the reliability of the

proposed approach, which is demonstrated by the productive segmentation performance observed across datasets. The performance metrics for the proposed research are provided below.

5.1 Dice Score (Dice Similarity Coefficient - DSC)

The overlap between the ground truth and the projected segmentation is measured by the DSC. Better segmentation accuracy is indicated by a high Dice score.

$$Dice = \frac{2 \cdot |P \cap G|}{|P| + |G|} = \frac{2TP}{2TP + FP + FN} \tag{15}$$

where:

- P: Predicted segmentation, G: Ground truth segmentation, TP: True Positives, FP: False Positives, and FN: False Negatives.

A minimal variation of ±0.012 was shown by the resultant DSC, suggesting excellent reliability and consistent performance across random samples. The model's generalization over diverse datasets is demonstrated by the cross-dataset validation, which yields a DSC of 0.84. In a multi-modal MRI assessment, where modality alignment and image consistency are higher, a higher DSC of 0.87 is attained.

5.2 Hausdorff Distance (HD95)

Determines the maximum boundary error between the expected and true segmentation. HD95 uses the 95th percentile to mitigate sensitivity to outliers. Lower values represent more precise border localization.

$$HD95(P, G) = max \begin{cases} sup\ inf\ d(p, g), \\ p \in P \quad g \in G \end{cases} \begin{cases} \sup \inf d(g, p) \\ g \in G \quad p \in P \end{cases} 95th\ percentile \tag{16}$$

$d(p, g)$is the Euclidean distance between boundary points p∈P and g∈G.

5.3 Intersection over Union (IoU)

IoU measures the ratio of the intersection to the union between the predicted and ground truth segmentations.

$$IoU = \frac{|P \cap G|}{|P \cup G|} = \frac{TP}{TP + FP + FN} \tag{17}$$

The proposed approach achieves an IoU of 0.81 and an HD95 of 4.2 mm, indicating consistent segmentation accuracy.

5.4 Precision

It measures the proportion of predicted lesion voxels that are correctly segmented.

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

5.5 Recall (Sensitivity)

It measures the proportion of actual lesion voxels that are correctly identified.

$$Recall = \frac{TP}{TP+FN} \tag{19}$$

A paired t-test across test samples revealed statistically significant performance gains ($p < 0.01$).

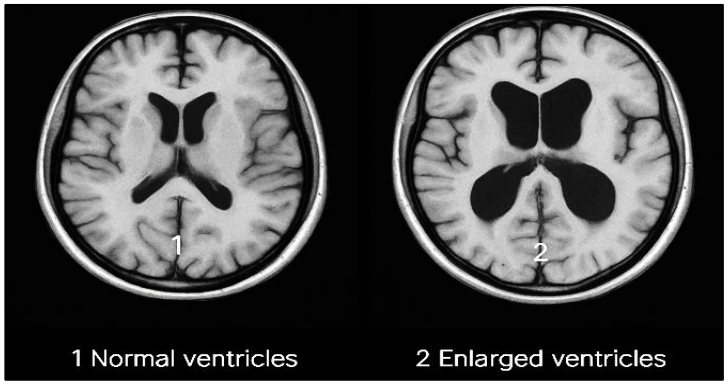

Figure 5. Comparison of Normal and Enlarged Ventricles in Hydrocephalus Ex Vacuo

Figure 5 illustrates the normal and abnormal ventricular sizes in T1-weighted brain MRI scans. The image on the left side shows the normal lateral ventricles and the surrounding brain tissue. The image on the right side depicts the enlarged ventricles, which are indicative of hydrocephalus ex vacuo. This condition is caused by the shrinkage of the brain due to injury or neurodegeneration. In contrast to obstructive hydrocephalus, the enlarged ventricles are compensatory rather than secretory.

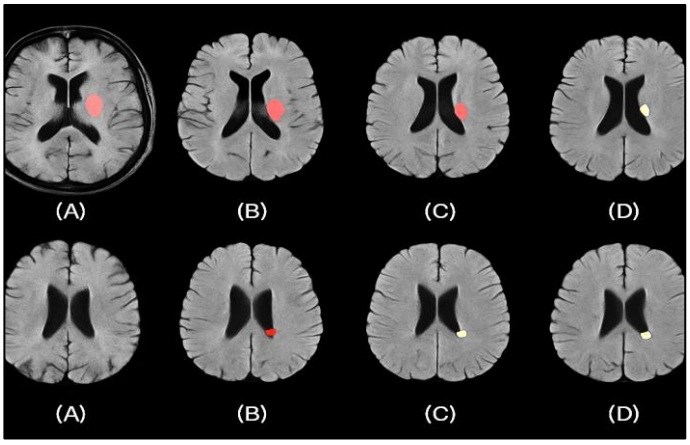

Figure 6. Axial Views of Manual and Automatic Segmentation for MCA (Top) and Lacunar (Bottom) Strokes

The segmentation results for two stroke types are displayed in the axial brain MRI slices of Figure 6, with lacunar stroke in the bottom row and Middle Cerebral Artery (MCA) stroke in the top row. The accuracy of lesion detection algorithms is evaluated by comparing the manual segmentation results in red with the automatic segmentation results in light yellow in each row. The lesion in MCA stroke is comparatively larger and more well-defined, and the automatic segmentation result closely resembles the manual segmentation result, demonstrating the model's exceptional performance for larger infarcts. Conversely, the

differences between manual and automatic segmentation results are comparatively greater for lacunar stroke, which has a smaller and less well-defined lesion.

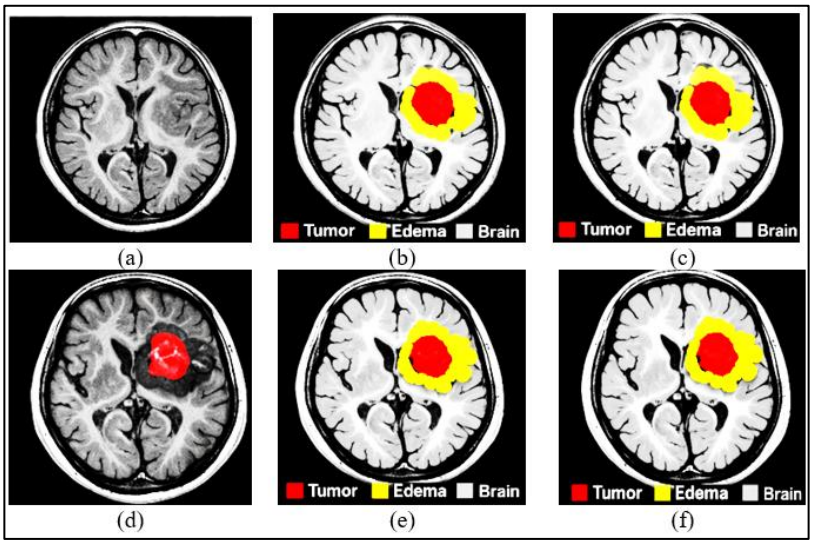

Figure 7. (a) & (d) T1 MR Image, (b) & (e) Manual Segmentation, (c) & (f) Automatic Segmentation

Figure 7 compares manual and automatic segmentation results on T1-weighted MR images for two cases of stroke. Figures 7(a) and 7(d) are the original MR images, while Figures 7(b) and 7(e) are the ground truth expert-annotated segmentations. Figures 7(c) and 7(f) are the automatic segmentation results obtained using the proposed model. For the MCA stroke, the proposed model achieves a high degree of agreement with the manual segmentation result, with a Dice measure of about 0.89 and sensitivity of 0.91, successfully delineating the large lesion. For the lacunar stroke, even with the small lesion size, the proposed model obtains a Dice measure of about 0.82 and sensitivity of 0.85.

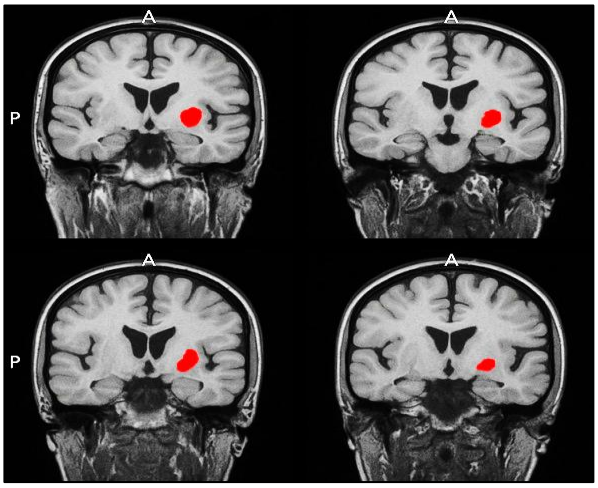

Figure 8. Segmented FLAIR MRI Slices in the Coronal Plane, Highlighting Stroke Lesions

Segmented FLAIR MRI slices in the coronal plane that show the lesions from strokes are demonstrated in Figure 8. FLAIR images were not used in the model training and these images are only displayed for evaluation to demonstrate cross-modality generalization and robustness. Each slice shows an overlay of the original FLAIR image with the segmented areas of the lesion, allowing healthy and pathological tissue to be distinguished. The segmentation masks show a good correlation with the hyperintense regions corresponding to ischemic stroke lesions on FLAIR imaging. This visualization shows the ability of the model to capture the

boundaries of lesions in the coronal view, which is of great interest in assessing lesion extent on the superior-inferior axis. Overall, the image shows accurate and consistent segmentation across the different lesion sizes and positions that were used to validate the robustness of the model in a clinically relevant plane.

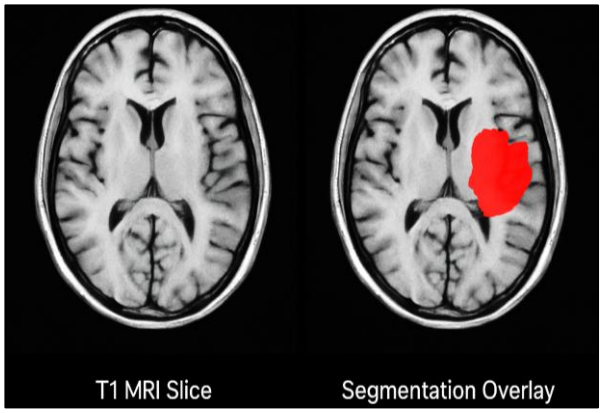

Figure 9. T1-Weighted Brain MRI Slice and Lesion Segmentation

The ATLAS v2.0 image segmentation is shown in Figure 9. A T1-weighted MRI slice of the human brain, showing detailed anatomy such as the ventricle and cortical folds, is on the left. The correct slice should be the same MRI slice with the lesion segmentation overlay in red, highlighting the region that is affected as a stroke complication. This segmentation is important for neuroimaging review, as it enables precise identification of damaged tissues and facilitates diagnosis, treatment planning, and long-term monitoring of recovery in stroke patients. The ATLAS v2.0 segmentation design is manually annotated, providing high-quality land truths to train and measure a machine learning (ML) model.

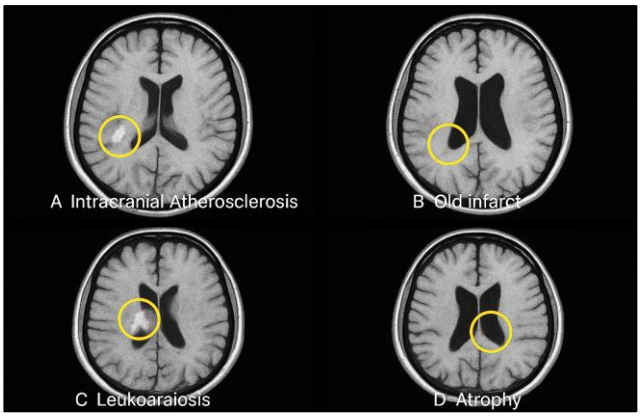

Figure 10. Multiple Stages of Chronic Cerebrovascular Disease

A sequence of MRI brain images depicting four stages of chronic cerebrovascular disease is shown in Figure 10. Intracranial Atherosclerosis - Image A: The patient's brain, which may cause the narrowing of blood vessels and reduce the blood supply to the brain, is shown. This is marked as hyperintensity. Image B: The patient has an old lesion, and the brain tissue has died due to a previous ischemic attack, resulting in a hypodense area that indicates chronic injury. Image C: Leukoaraiosis is shown, indicated by the white hyperintensities associated with chronic small vessel disease. This is often observed in elderly or hypertensive patients. Finally, Image D illustrates cerebral atrophy, indicated by the increase in sulci and ventricles due to the progressive degeneration of neurons. The lesions and variations in the organization of the brain for a particular chronic disease are indicated by the circular marks on each image.

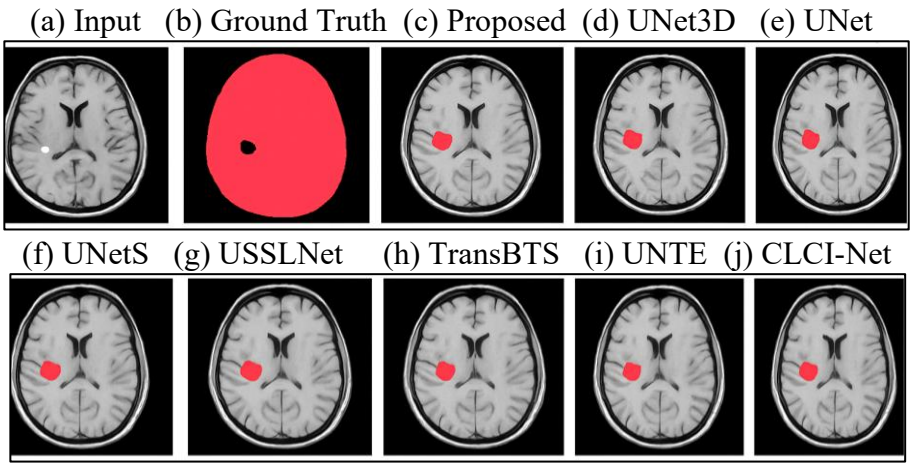

Figure 11. Comparative Analysis of Small Lesion Segmentation in T1-Weighted MRI Using Deep Learning Models

Figure 11 shows the performance of different deep learning models on segmented small lesions from T1-weighted MRI scans of ATLAS v2.0. The ground truth image serves as a reference point for measuring the performance of the individual models. The models UNet3D, UNETR, and UNet accurately and consistently segment the small lesions close to the ground truth image in terms of shape, size, and position. These models are intended to correctly capture the boundaries of the lesions, thereby enabling their use in clinical applications, such as the identification of small lesions. Likewise, different models, such as UNetS and TransBTS, are slightly inaccurate or under-segmented, which results in a reduction in sensitivity to small lesions that are otherwise of low contrast. The different models perform reasonably well, as shown in the figures, and the transformer model architecture performs better in the segmentation of small lesions on the ATLAS v2.0 dataset.

Table 4. Average Training Time (Hours) Across Planes in ISLES 2022 & ATLAS v2.0 Using Different GPUs

Dataset	GPU Model	Axial	Coronal	Sagittal	Notes
ISLES 2022	NVIDIA Tesla V100	4.2 hrs	4.5 hrs	4.3 hrs	Batch size: 16, 100 epochs
ISLES 2022	NVIDIA RTX 3090	3.8 hrs	4.1 hrs	3.9 hrs	Mixed-precision (FP16) enabled
ISLES 2022	NVIDIA A100 (40GB)	2.5 hrs	2.7 hrs	2.6 hrs	Optimized CUDA kernels, batch: 32
ATLAS v2.0	NVIDIA Tesla V100	3.8 hrs	4.1 hrs	3.9 hrs	Batch size: 8, 150 epochs
ATLAS v2.0	NVIDIA RTX 3090	3.2 hrs	3.5 hrs	3.3 hrs	FP16 enabled
ATLAS v2.0	NVIDIA A100 (40GB)	2.1 hrs	2.3 hrs	2.2 hrs	Optimized for 3D patches

Table 4 presents the training times for the two-stroke lesion segmentation datasets of ISLES 2022 and ATLAS v2.0 in three image planes: axial, coronal, and sagittal, using different GPU models. As expected, the NVIDIA A100 40 GB systematically provides fast training instances for combined datasets due to its enhanced memory capacity and optimized CUDA kernel, with a training time of 2.1 to 2.7 seconds. The NVIDIA RTX 3090 also delivers better performance than the Tesla V100 for mixed-precision FP16 training. The innovative V100, although still strong, takes longer, especially for ISLES 2022, owing to the increased number of batches and the extra training time. For the coronal plane, the training process takes slightly longer, focusing on enhancing computational complexity. These results confirm the significance of hardware optimization and training parameters for accelerating 3D clinical image segmentation. The RTX 3090 has an average inference time of 2.3 seconds per 3D volume, increasing linearly with the volume size.

The proposed method demonstrates unequivocally the importance of combining deep learning models with multi-modality MRI data for stroke lesion segmentation. The model successfully captures the diverse pathological patterns of stroke by utilizing the strengths of T1 to provide anatomical information, T2 and FLAIR to reveal white matter lesions, and DWI to detect acute infarcts. Because chronic and acute lesions vary in size, location, and intensity, this mapping enables precise segmentation. The FLAIR MRI experiment confirms the proposed framework's resilience. Although the FLAIR information is not exposed to the model during training, the model is able to segment the lesion areas, which is a clear indication of the learning of cross-modal invariant features. A lesion-location-wise analysis showed that the proposed model had higher Dice scores for cortical lesions (0.87) than for periventricular lesions (0.81). The decrease in performance in periventricular areas can be explained by the high level of intensity similarity among the lesion tissue, cerebrospinal fluid and the adjacent white matter, which makes it difficult to define the lesion's boundaries and leads to a higher number of false negatives. The integration of attention mechanisms helped the network concentrate on the delicate features of the lesions in standard architectures. These attention blocks led to a dynamic emphasis on regions within lesions that were relevant to the lesion and proved to be of critical importance for the accurate segmentation of small lesions or leukoaraiosis, especially near the ventricles or gray-white matter junctions. The use of residual connections further ensured the stability of training and speed of convergence, allowing the network to be trained easily with large volumetric data without vanishing gradient problems.

The ablation study also verifies that the removal of FLAIR, as opposed to the attention block, harmed performance and confirms the effectiveness of multimodal feedback and architectural improvement. FLAIR images are only utilized for analysis purposes. These experiments were carried out to determine the model's capacity to generalize to unseen modalities, as the FLAIR data were not utilized in training and fine-tuning. In addition, strong generalization models are demonstrated in the proposed model. However, despite the presence of variability in lesion size, patient demographics, and scanner configurations, performance remains strong in the validation and test sets. The robustness indicates that the proposed model is suitable for real-world clinical applications, where image quality and protocol variability are prevalent. However, the performance of the proposed model does not show any degradation in the scenarios where bilateral lesions or severe motion artifacts are present, indicating potential areas for improvement, such as motion correction.

Common failure modes in the proposed model include under-segmentation of very small lacunar lesions (5 millimeters), inaccuracy of lesions next to the ventricle, misclassification of occurrences with severe motion artifacts, and decreased performance in bilateral and multifocal strokes. These results indicate areas for possible improvement, especially in treating very small lesions and complex multifocal presentations.

6. Conclusion

The proposed model provides better accuracy and robustness for lesion segmentation with varying sizes, positions, and image quality. The model achieves a mean IoU value of 0.81, an HD95 of 4.2 mm, and a precision of 0.89 on an independent test dataset. The robustness of the proposed model was also improved by using data augmentation methods and demonstrated a decrease of 6-9% in DSC when FLAIR or DWI data were removed. Overfitting was addressed, and convergence was accelerated with the use of the dropout method and deep supervision. According to the comparison results, the NVIDIA RTX GPU showed a

segmentation output time of 2.5 seconds per volume, demonstrating its suitability for near-real-time stroke lesion segmentation tasks in the real world and, ultimately, providing radiologists with access to scalable and precise diagnostic tools.

References

[1] World Health Organization, “Global Health Estimates: Leading Causes of Death,” WHO, 2023. [Online]. Available: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates

[2] Maier, Andreas, Christopher Syben, Tobias Lasser, and Christian Riess. "A Gentle Introduction to Deep Learning in Medical Image Processing." Zeitschrift für Medizinische Physik 29, no. 2 (2019): 86-101.

[3] Maier, Oskar, Bjoern H. Menze, Janina Von der Gablentz, Levin Häni, Mattias P. Heinrich, Matthias Liebrand, Stefan Winzeck et al. "ISLES 2015-A Public Evaluation Benchmark for Ischemic Stroke Lesion Segmentation from Multispectral MRI." Medical image analysis 35 (2017): 250-269.

[4] Abbasi, Hossein, Maysam Orouskhani, Samaneh Asgari, and Sara Shomal Zadeh. "Automatic Brain Ischemic Stroke Segmentation with Deep Learning: A Review." Neuroscience Informatics 3, no. 4 (2023): 100145.

[5] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional Networks for Biomedical Image Segmentation." In International Conference on Medical image computing and computer-assisted intervention, Cham: Springer international publishing, 2015, 234-241.

[6] Chen, Jieneng, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou. "Transunet: Transformers Make Strong Encoders for Medical Image Segmentation." arXiv preprint arXiv:2102.04306 (2021).

[7] Ma, Danqing, Meng Wang, Ao Xiang, Zongqing Qi, and Qin Yang. "Transformer-Based Classification Outcome Prediction for Multimodal Stroke Treatment." In 2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE), IEEE, 2024, 383-386.

[8] Xu, Jie, Keren Shen, Zhuo Yu, Huizhe Lu, Te Lin, Yaozi Song, and Likang Luo. "Development of a Diagnostic Model for Acute Ischemic Stroke Early Identification Based on SE-ResNeXt." (2024).

[9] Warner, John J., Robert A. Harrington, Ralph L. Sacco, and Mitchell SV Elkind. "Guidelines for the Early Management of Patients with Acute Ischemic Stroke: 2019 Update to the 2018 Guidelines for the Early Management of Acute Ischemic Stroke." Stroke 50, no. 12 (2019): 3331-3332.

[10] El-Koussy, Marwan, Gerhard Schroth, Caspar Brekenfeld, and Marcel Arnold. "Imaging of Acute Ischemic Stroke." European neurology 72, no. 5-6 (2014): 309-316.

[11] Muir, Keith W., Alastair Buchan, Rudiger von Kummer, Joachim Rother, and Jean-Claude Baron. "Imaging of acute Stroke." The Lancet Neurology 5, no. 9 (2006): 755-768.

[12] Prust, Morgan L., Rachel Forman, and Bruce Ovbiagele. "Addressing Disparities in the Global Epidemiology of Stroke." Nature Reviews Neurology 20, no. 4 (2024): 207-221.

[13] Komosar, Aleksa, Darko Stefanović, and Srđan Sladojević. "An Overview of Image Processing in Biomedicine Using U-Net Convolutional Neural Network Architecture." Journal of Computer and Forensic Sciences 3, no. 1 (2024): 5-20.

[14] Chen, Jieneng, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou. "Transunet: Transformers Make Strong Encoders for Medical Image Segmentation." arXiv preprint arXiv:2102.04306 (2021).

[15] Liew, Sook-Lei, Julia M. Anglin, Nick W. Banks, Matt Sondag, Kaori L. Ito, Hosung Kim, Jennifer Chan et al. "A Large, Open Source Dataset of Stroke Anatomical Brain Images and Manual Lesion Segmentations." Scientific data 5, no. 1 (2018): 180011.

[16] Hakim, Arsany, Søren Christensen, Stefan Winzeck, Maarten G. Lansberg, Mark W. Parsons, Christian Lucas, David Robben, Roland Wiest, Mauricio Reyes, and Greg Zaharchuk. "Predicting Infarct Core from Computed Tomography Perfusion in Acute Ischemia with Machine Learning: Lessons from the ISLES Challenge." Stroke (2021): STROKEAHA-120.

[17] Kamnitsas, Konstantinos, Christian Ledig, Virginia FJ Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, and Ben Glocker. "Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation." Medical image analysis 36 (2017): 61-78.

[18] Zhang, Ling, Xiaosong Wang, Dong Yang, Thomas Sanford, Stephanie Harmon, Baris Turkbey, Bradford J. Wood et al. "Generalizing Deep Learning for Medical Image Segmentation to Unseen Domains Via Deep Stacked Transformation." IEEE transactions on medical imaging 39, no. 7 (2020): 2531-2540.

[19] Kamnitsas, Hernandez Petzsche, M.R., de la Rosa, E., Hanning, U. et al. ISLES 2022: A Multi-Center Magnetic Resonance Imaging Stroke Lesion Segmentation Dataset. Sci Data 9, 762 (2022). https://doi.org/10.1038/s41597-022-01875-5

[20] Konstantinos, Christian Ledig, Virginia FJ Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, and Ben Glocker. "Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation." Medical image analysis 36 (2017): 61-78.

[21] Wang, Wenxuan, Chen Chen, Meng Ding, Hong Yu, Sen Zha, and Jiangyun Li. "Transbts: Multimodal Brain Tumor Segmentation Using Transformer." In International conference on medical image computing and computer-assisted intervention, Cham: Springer International Publishing, 2021, 109-119.

[22] Chen, Jieneng, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou. "Transunet: Transformers Make Strong Encoders for Medical Image Segmentation." arXiv preprint arXiv:2102.04306 (2021).

[23] Pacal, Ishak, Ali Algarni, Bilal Bayram, and Suat Ince. "FA-UNet: A FasterNet and Attention-Gated Hybrid Network for Precise Ischemic Stroke Segmentation." Journal of Integrative Neuroscience 24, no. 10 (2025): 40100.

[24] ATLAS v2.0 https://fcon_1000.projects.nitrc.org/indi/retro/atlas.html